

RESEARCH ARTICLE

Applied Cellular Physiology and Metabolic Engineering

Examining organic acid production potential and growth-coupled strategies in *Issatchenkia orientalis* using constraint-based modeling

Patrick F. Suthers^{1,2} | Costas D. Maranas^{1,2} 

¹Department of Chemical Engineering, The Pennsylvania State University, University Park, Pennsylvania, USA

²Center for Advanced Bioenergy and Bioproducts Innovation, The Pennsylvania State University, University Park, Pennsylvania, USA

Correspondence

Costas D. Maranas, Department of Chemical Engineering, The Pennsylvania State University, University Park, PA 16802, USA.
Email: costas@psu.edu

Funding information

US Department of Energy, Grant/Award Numbers: DE-SC0018260, DE-SC0018420

Abstract

Growth-coupling product formation can facilitate strain stability by aligning industrial objectives with biological fitness. Organic acids make up many building block chemicals that can be produced from sugars obtainable from renewable biomass. *Issatchenkia orientalis* is a yeast strain tolerant to acidic conditions and is thus a promising host for industrial production of organic acids. Here, we use constraint-based methods to assess the potential of computationally designing growth-coupled production strains for *I. orientalis* that produce 22 different organic acids under aerobic or microaerobic conditions. We explore native and engineered pathways using glucose or xylose as the carbon substrates as proxy constituents of hydrolyzed biomass. We identified growth-coupled production strategies for 37 of the substrate-product pairs, with 15 pairs achieving production for any growth rate. We systematically assess the strain design solutions and categorize the underlying principles involved.

KEYWORDS

constraint-based modeling, metabolic engineering, nonmodel yeast

1 | INTRODUCTION

Metabolic engineering approaches modify cellular activities in order to improve the production of metabolite or protein products.¹ The rise of genomic sequencing tools has enabled the rapid reconstruction of genome-scale metabolic models for a number of organisms.² These models can be used to inform intervention strategies through the use of constraint-based reconstruction and analysis (COBRA) approaches.³ Current advances in constraint and machine learning-based metabolic modeling have been recently reviewed.⁴ The use of such approaches has aided the successful commercialization of processes to produce 1,4-butanediol,⁵ lactic acid,^{6,7} and heterologous proteins for therapeutic use as biopharmaceuticals,⁸ among others.

One possible approach for strain design is to couple the production of a desired product to the growth of a microbe.⁹ This alignment

of industrial objectives with biological fitness ones can improve pathway stability by reducing the pressure on selection to divert carbon away from the product toward biomass. Such a designed strain can then be acted upon adaptive evolutionary strategies¹⁰ to indirectly select for improved product yield through growth selections.¹¹ The foundational computational tool for rational strain design that generates growth-coupled production (GCP) strategies is OptKnock.¹² OptKnock solves a bilevel optimization problem to pinpoint a set of reactions that should be eliminated simultaneously from a metabolic network in order to ensure that the desired product becomes a potential byproduct of biomass formation. Each reaction elimination (RE) design solution can be examined by making the indicated changes on bounds to the reactions and then plotting a metabolic production envelope¹³ that projects the accessible flux space onto the plane of growth rate and the target's production rate. Through the use of

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. *Biotechnology Progress* published by Wiley Periodicals LLC on behalf of American Institute of Chemical Engineers.

integer cuts, alternative solutions can be computed. These alternative solutions can be examined for implementation via the gene-protein-reaction (GPR) associations that are part of genome-scale metabolic models, as well as subjected to additional analyses and rankings, by using Flux Balance Analysis (FBA).¹⁴ Extensions and improvements to OptKnock has been a fertile area of algorithmic research and has been reviewed.¹⁵ These tools include examples such as RobustKnock,¹⁶ FOCAL,¹⁷ ReacKnock,¹⁸ and SMET.¹⁹

When examining generated RE design solution sets, a number of classifications and production phenotype metrics can be helpful to describe each solution that produces product p and biomass X from substrate s , as shown by production envelopes in Figure 1. For simplicity, in this section, we make substrate notations implied. Qualitatively, a solution can be partitioned broadly into various cases based on the production phenotype.^{12,20–22} Here, we adopt an existing classification scheme²² with some minor extensions and distinguish five cases: null, potentially, weakly, and directionally growth-coupled production (GCP) and substrate-uptake-coupled production; we further use a collective category termed strongly growth-coupled production comprising directionally growth-coupled production and substrate-uptake-coupled production. The first, null (referred to hereafter as \emptyset GCP), indicates no growth coupling and occurs when at the maximum biomass production rate, $g_{\max} := v_{X,\max}$, the maximum product rate, $v_{p,\max}^g$, is zero. Typically, the starting network will be \emptyset GCP. For the second, potentially growth-coupled production (pGCP), the product rate, v_p , displays a phenotype consisting of equivalent optimal solutions that does not ensure production of the desired compound. That is, at the maximum biomass production, g_{\max} , the maximum product rate $v_{p,\max}^g$ is positive whereas the minimal product rate, $v_{p,\min}^g$, is

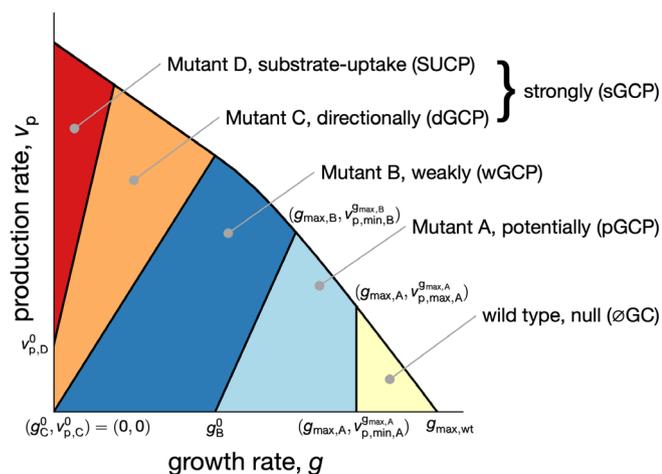


FIGURE 1 Illustration of production envelopes for a wild type strain and four mutant strains having different qualities of growth-coupled production. The accessible solution space is below each curve and the production envelopes are shown stacked with each including any regions to the left. For example, the production envelope of mutant strain C encompasses the regions marked C and D, as mutant strain D is drawn on top of mutant strain C. The wild type strain produces no product at its maximum growth rate and thus has null growth coupling. Key product and growth rates used in defining qualities are indicated.

zero. For weakly and strongly growth-coupled production, conversely, instead the solution displays a phenotype with $v_{p,\min}^g > 0$. For weakly growth-coupled production (wGCP), the production envelope allows for zero production of the target (i.e., $v_{p,\min} = 0$) until reaching some positive growth rate, g^0 , after which its production is always greater than zero. For strongly growth-coupled production (sGCP, i.e., both directionally growth-coupled production and substrate-uptake-coupled production) the production envelope displays positive production of the target for all growth rates greater than zero (i.e., $v_{p,\min}^g > 0$ for all $g > 0$) and so thus has a maximum growth at production onset, $g^0 := v_{X,\max}^{v_{p,\min}^g=0}$, of zero. Directionally growth-coupled production (dGCP) and substrate-uptake-coupled production (SUCP) differ in the minimal production rate at no growth, $v_{p,\min}^0 := v_{p,\min}^{g=0}$. dGCP has $v_{p,\min}^0 = 0$, indicating that any growth necessitates product formation, whereas SUCP has $v_{p,\min}^0 > 0$, indicating that product is always produced, even at no growth. For some wGCP or sGCP cases, instead of a singular product rate $v_p^g = v_{p,\max}^g = v_{p,\min}^g$ there exist equivalent optimal solutions (i.e., $v_{p,\max}^g \geq v_p^g \geq v_{p,\min}^g > 0$). The RE design solutions can be ranked by quantitative criteria such as converting the above-mentioned product and biomass production rates into equivalent product yields (Y), substrate-specific productivity (SSP),²⁰ growth-coupling strength (GCS).²¹

An industrially viable process relying on a microbial production of organic acids needs to be an efficient producer (i.e., have a high yield from sugar).²³ Although some of the above-mentioned organic acids have been produced using production hosts such as *Escherichia coli* (e.g., succinic acid), the industrial processes require pH neutralization and thereby result in byproducts such as gypsum. Thus, ideally a production host tolerates the low pH associated with a high titer²³ which can enable product separation without extensive neutralization. The yeast *Issatchenkia orientalis* (also known as *Pichia kudriavzevii*, *Candida glycerinogenes*, and *Candida krusei*²⁴) has been proposed to be one such candidate host since it exhibits tolerance to high levels of succinic acid, itaconic acid, adipic acid, and acetic acid.²⁵ Strains can produce ethanol in media containing 5% sodium sulfate at pH 2.0.²⁶ Recombinant *I. orientalis* strains can produce titers of 11.6 g/L succinic acid²⁷ and 15–20 g/L lactic acid under anaerobic conditions in an unbuffered medium at a pH of 2.²⁸ Indeed, a strain was reported to produce as much as 154 g/L D-lactic acid at a pH of 4.7 after genetic modifications and subsequent adaptations to high lactic acid concentrations.²⁹ *I. orientalis* is capable of growth on a number of carbon substrates, with growth/no growth assays previously tested on 34 carbon substrates²⁶ as well as in a separate study of 26 carbon substrates.³⁰ In the former, seven of the carbon substrates scored positive (i.e., glucose, lactose, glycerol, lactic acid, succinic acid, citric acid, and ethanol) and four were delayed positive (i.e., xylose, sucrose, xylitol, and glucono-1,5-lactone).²⁶ In the latter, six scored positive (i.e., same as the positives in the former excepting lactose).³⁰ Quantitative analysis of growth rates was recently evaluated for glucose, glycerol, fructose, succinic acid, lactic acid, citric acid, and ethanol.³¹ Isolated strains have been found to grow at a pH of 2.5 on hemicellulosic and cellulosic oligosaccharides obtained by two-step extraction with sulfuric acid from six plant sources.³²

Recent advances in genetic systems for this nonmodel microorganism^{33,34} expedite its domestication as an industrial host, and a recently published genome-scale metabolic (GSM) model³¹ allows for the application of COBRA³ tools. In the current work, we extend our limited-scope examination of succinic acid production³¹ and employ constraint-based modeling to exhaustively examine the potential maximum theoretical yields and dependence of yield on oxygen uptake for a bevy of 22 organic acids from three to six carbons, including seven of the organic acid DOE building-blocks, for individually both glucose and xylose carbon substrates. For non-native organic acids, we introduce synthetic pathways that enable their production in the GSM model. We then use OptKnock to perform a large-scale computational study to identify RE solutions that facilitate GCP strain designs. For each solution, we classify the various GCP strain designs with qualitative (i.e., pGCP, wGCP, dGCP, or SUCP) and quantitative (i.e., $Y_{P/S,\min}^{g,\max}$, $Y_{P/S,\max}^{g,\max}$, SSP, Y^O , g^O , and GCS) characteristics. We examine the distribution of RE solutions that occur for multiple organic acids. We also examine the impact on introducing other reactions into the system that can increase carbon yield for specific targets.

2 | MATERIALS AND METHODS

2.1 | Databases and identifiers

We used BiGG Models,³⁵ MetaNetX,³⁶ ModelSEED,³⁷ ChEBI,³⁸ and PubChem³⁹ for information about the metabolites and reactions, including formula and charge information for the selected organic acids. In general, we followed BiGG nomenclature for metabolite id, with a few notable exceptions. Propionate has two entries in BiGG: ppa and prpnte; we use the former because it has wider usage within the database. Malonate likewise has two entries in BiGG: HC00319 and malon. We noted that the latter entry did not contain the formula or charge information but did contain links to other databases, whereas the former did have formula and charge, but no database links despite being used in more models. Thus, we merged the information from both entries into the id malon and used that herein, since it more closely follows the typical naming scheme within BiGG. For 3-hydroxypentanoate, we used 3hpt (unassigned within BiGG) instead of its BiGG id R_3hpt to lessen the chance of confusion with a reaction id. Metabolites that do not occur in BiGG were given ids which are unassigned within BiGG. We used MetaNetX version 4.2 during this process to establish metabolite and reaction properties and annotations.

2.2 | Modeling simulations

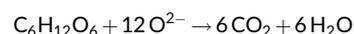
Flux balance analysis (FBA) was used throughout the process for model predictions.¹⁴ The genome-scale metabolic model used for *I. orientalis* SD108 is *ilsor850*.³¹ For comparisons for carbon substrates, molar amounts of each carbon substrate uptake rate were set to 10.0 mmol gDW⁻¹ h⁻¹ glucose or xylose, respectively, during simulations. Oxygen uptake rate was limited to no more than 18.18 mmol gDW⁻¹ h⁻¹ during all simulations unless otherwise noted, as this is

the minimal oxygen uptake that does not impact the maximum growth rate on a 10 mmol glucose gDW⁻¹ h⁻¹ basis. Model simulations of growth phenotype or product formation were obtained using FBA with the objective of maximizing the flux of the biomass reaction (v_{biom}) or product exchange flux (v_{EX_p}) corresponding to the target product. Unless otherwise indicated, a minimal biomass reaction flux was set as 10% of the maximum biomass reaction flux found under the uptake conditions using FBA. For all cases, we used the value of 1.0 for nongrowth-associated ATP maintenance constraint that was set during the model's reconstruction process.³¹

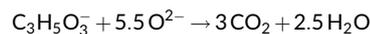
During initial stages, we examined if the selected product natively occurred in the wild type network. If not, then we added a pathway that enabled its production. The added metabolites and reactions are provided in SBML format in Supporting Information, Data S1, and the accompanying notes field of each reaction therein includes the associated product(s) for which it is added to the model. Using constraints corresponding to aerobic minimal media conditions with the appropriate carbon substrate, we examined coupled reaction sets using the Flux Coupling Finder (FCF).⁴⁰ For any two fluxes in a fully coupled reaction set, a nonzero flux for a given one implies a fixed nonzero value for the other member, and vice versa. A gene knockout was translated to the corresponding reaction elimination(s) by examining the Boolean GPR associations; each indicated reaction was appropriately eliminated in the model by setting the corresponding upper and lower flux bounds both to zero. A gene or reaction was classified as essential if the maximal growth rate of the corresponding knockout mutant or reaction elimination was calculated by FBA to be less than 0.001 h⁻¹.

2.3 | Theoretical analysis of production potential

The production potential of each substrate and product pair was determined using both chemical and biochemical analyses in context of *I. orientalis* using FBA. For the chemical analysis, we used two methods. The first tabulates the ratio of the moles of each product that could be produced per mole of the sugar carbon substrate based solely on the number of carbons present in each compound, which we designate the carbon yield, Y_{P_c/S_c} . The second, which we designate the available electron yield, Y_{P_e/S_e} , uses metabolite charges and molecular formulae to calculate the available electrons for each product and the moles of molecular oxygen required for complete combustion. For instance, the reaction for the complete combustion of glucose is



which results in $0 + 2 \times 12 = 24$ available electrons for glucose; for 3-hydroxypropionic acid (which has a charge of -1 under the physiological conditions in the model) it is:



which results in $1 + 2 \times 5.5 = 12$ available electrons for 3-hydroxypropionic acid. Thus, 3-hydroxypropionic acid has a theoretical yield $Y_{3\text{hpp}_e/\text{glc}_e}$ of $24/12 = 2$ mol/mol glucose. For succinic

acid, similar computations result in $2 + 2 \times 6 = 14$ available electrons for a $Y_{\text{succ}_e/\text{gl}_c}$ of $24/14 = 1.71$ mol/mol glucose. Since all the products and substrates examined herein contain only carbon, hydrogen, and oxygen, we generalized the equation for available electrons of a compound with a chemical formula of $C_xH_yO_z$ as $e = 4x + y - 2z - c$, where c is the net charge. These computations assume no additional reducing power from a source such as H_2 .

In order to determine the biologically achievable theoretical maximum, the workflow was to 1) include any non-native pathways as required, 2) set the indicated carbon substrate and oxygen uptake constraints, as given above, 3) use FBA to maximize the exchange flux corresponding to the target product, and 4) store the resulting objective and flux values $v_{j,s}$. The biological theoretical yield potential is then computed as $Y_{P,S} = v_{p,s,\text{max}}/v_s$.

2.4 | Reaction subsets and model preprocessing

We determined a reduced set of reactions upon which the OptKnock algorithm operates for each substrate and product pair, largely following and extending a procedure previously outlined.²⁰ Specifically, we defined and populated sets of reactions (J) for each substrate-product pair as follows. Using the appropriate model for each pair (i.e., containing the base model and only the required non-native pathway that enabled product formation) we used Flux Variability Analysis (FVA)⁴¹ to find the blocked reactions that cannot carry any flux (J^{blocked}) for the medium conditions. By examining the Systems Biology Ontology (SBO)⁴² annotations in the model files, we identified all transport reactions ($J^{\text{transport}}$) and exchange reactions (J^{exchange}). By examining the GPR associations for genes G in the model file, we identified all reactions without a listed GPR (J^{nonGPR}). Essential reactions ($J^{\text{essential}}$) and genes ($G^{\text{essential}}$) were determined by FBA. $G^{\text{essential}}$ and GPR associations were used to define reactions with no way to eliminate them other than turning off an essential gene ($J^{\text{inaccessible_by_GPR}}$). The coupled reaction sets found using FCF were used to define (J^{coupled}) that contained all but one reaction for each set of equivalent reaction eliminations; the excluded reaction was the first alphabetically in each set of coupled reactions that was not a member of $J^{\text{transport}} \cup J^{\text{nonGPR}} \cup J^{\text{inaccessible_by_GPR}}$. We defined reactions accessible for elimination as $J^{\text{accessible}} = J - (J^{\text{blocked}} \cup J^{\text{transport}} \cup J^{\text{nonGPR}} \cup J^{\text{inaccessible_by_GPR}} \cup J^{\text{coupled}})$.

We identified reactions for which a non-zero flux directly impacts product yield negatively. Based on convention from OptKnock, we added the constraint $v_{X,S} \geq 0.1v_{X,S,\text{max}}$ (i.e., the minimum growth constraint), which represents an approximation of the minimum growth required to obtain the necessary cell mass in a bioreactor in a reasonable timescale. For notational convenience, we define $g := v_{X,S}$. Then we used FBA to compute the resulting product yield at this constraint, $Y_{P/S,\text{max}}^{g=0.1g_{\text{max}}}$. Next, we fixed the lower bound on the product's corresponding exchange flux to $Y_{P/S,\text{max}}^{g=0.1g_{\text{max}}}$, performed FVA over set $J^{\text{accessible}}$, and found any reactions j for which $v_{j,\text{min}}^{g=0.1g_{\text{max}}} \leq 0 \leq v_{j,\text{max}}^{g=0.1g_{\text{max}}}$. For some, zero was the only permitted value (i.e., $v_{j,\text{min}}^{g=0.1g_{\text{max}}} = 0 = v_{j,\text{max}}^{g=0.1g_{\text{max}}}$) and any nonzero flux through them would negatively impact product yield); we placed these reactions in set

$J^{\text{must_off}}$. Others had non-zero $v_{p,\text{min}}^{g=0.1g_{\text{max}}}$ and/or $v_{j,\text{max}}^{g=0.1g_{\text{max}}}$; we placed these reactions in set $J^{\text{may_off}}$. Conversely, in order to identify reactions for which a zero flux directly negatively impacts product yield, for all reactions j in $J^{\text{accessible}}$ we set both upper and lower bounds of said reaction to zero (i.e., eliminated the reaction) and ran the optimization problem to maximize product yield, $v_{p,s}^{v_j=0}$. Reactions for which this analysis resulted in $Y_{P/S}^{v_j=0} < Y_{P/S}$ were examined and those with ratio $Y_{P/S}^{v_j=0}/Y_{P/S} < 0.90$ were included in set $J^{\text{should_on}}$.

For each substrate-product pair, we used OptKnock to generate M strain designs made up of k simultaneous reaction eliminations. We enforced a minimal growth rate constraint of 10% of the maximum for the starting model, as mentioned above, as well as a minimal product rate constraint corresponding to 10% carbon yield. OptKnock was implemented such that it could resume computations given a set of previous solutions. This implementation also enables parallelization of computations for substrate/product pair in two by fixing one or more groups of binary variables. For instance, this procedure allowed forcing reactions such as mitochondrial ATP synthase, v_{ATPS_m} , on or off, as well as fixing individual reactions in $J^{\text{must_off}}$ or $J^{\text{may_off}}$ to off or in $J^{\text{should_on}}$ to on (or combinations thereof). It could also be used for examination of higher order sets by fixing off reactions in that were observed to occur in all solution sets or in one or a particular set. For solution sets up to size 5 RE, we ran OptKnock only using $J^{\text{accessible}}$ and exhaustively examined the solution space. That is, we either did not fix any reactions to on or off (excluding integer cuts) or we ran parallel cases fixing the reaction(s) both on and off (or combinations thereof) and subsequently merged the results. For higher order sets, we used variable fixing to make OptKnock computationally tractable. We subsequently used FVA to sort non-unique solutions N that had exchange fluxes with lower bounds of zero at the maximum biomass flux (i.e., $v_{p,\text{min}}^{g_{\text{max}}} = 0$; pGCP); the remaining unique solutions U were ones for which the product was guaranteed to be produced at non-zero amounts at maximum biomass (i.e., wGCP or sGCP). We then define sets of reactions in each such solution category as $J_{p,s,U}^{\text{OK}}$ or $J_{p,s,U}^{\text{OK}}$, the latter of which can be further subdivided into wGCP, dGCP, and SUCP: $J_{p,s,wGCP}^{\text{OK}}$, $J_{p,s,dGCP}^{\text{OK}}$, and $J_{p,s,SUCP}^{\text{OK}}$ respectively.

We examined the feasibility of implementing RE designs using gene knockouts by using the GPR associations for each reaction in the design. For reactions with GPR associations containing only AND operators, we selected the first gene listed for knockout. For those reactions containing only OR operators, we selected all genes listed for knockout. For those reactions containing both AND and OR operators, we selected the first single gene listed that when knocked out would ensure the reaction was eliminated. We then knocked out all selected genes simultaneously and computed the associated production envelope, which we then compared with the production envelope for the corresponding RE design.

2.5 | Metrics computed for strain designs

For each growth-coupled design solution m of the substrate-product pair we compute the metrics for biological product yields $Y_{P/S,m}$,

namely $Y_{P/S,max}$, $Y_{P/S,max}^{g_{max}}$, $Y_{P/S,min}^{g_{max}}$, $Y_{P/S,min}^0$, substrate-specific productivity ($SSP_{p,s,m}$)²⁰ and growth-coupling strength ($GCS_{p,s,m}$)²¹. SSP has also been called the biomass–product coupled yield (BPCY).⁴³ These quantities are calculated as the following. Each $Y_{P/S,m}$ is computed as described above for the biological theoretical maximum from the corresponding product formation, $v_{p,s}$, and substrate uptake, v_s . For instance, from the maximum product formation, $v_{p,s,max}$, we can calculate

$$Y_{P/S,max,m} = \frac{v_{p,s,max,m}}{-v_s}$$

with negative v_s used in the denominator to make the value positive, since uptake fluxes have a negative sign by convention. Similarly, we compute $Y_{P/S,min}^{g_{max}}$ and $Y_{P/S,max}^{g_{max}}$ using FVA on the product exchange flux when constraining the biomass flux to $g_{max,m}$. Substrate-specific productivity is computed from the product yield, $Y_{P/S,m}$, and growth rate, g . Our SSP metric is the minimal productivity found at the maximum biomass produced for the given constraints, $g_{max,m}$, which we normalized relative to the wild type and carbon yield thus

$$SSP_{p,s,m} = \frac{Y_{P/S,min,m} \cdot g_{max,m}}{Y_{P_c/S_c} \cdot g_{max}} = \frac{Y_{P/S,min,m}}{Y_{P_c/S_c}} \cdot \frac{v_{X,s,max,m}}{v_{X,s,max}}$$

resulting in a number between 0 and 1. The boundary for SSP for a given mutant m can be found by instead evaluating SSP along all allowable values $Y_{P/S,max,m}$ along the growth range $0 \leq g \leq g_{max,m}$ and substituting these values in the numerators in the expression above. Finally, GCS is a measure that takes into account the total area, TA_m , of the production envelope beneath the maximal production rate of the RE strain m , and the inaccessible area, IA_m , of the production envelope beneath the minimal production rate of the RE strain m .²¹ The GCS metric also considers the minimally guaranteed target product yield at maximal growth $Y_{P/S,max,m}^{g_{max}}$ and the biologically maximum theoretical yield $Y_{P/S,max}$ for the substrate–product pair in the form of a penalty for reduction in yield. The piecewise function also uses the maximum growth rate for which minimal production of the selected product is zero, $v_{X,s,min}^{v_{p,s,min}=0}$, and the minimal product rate at no growth, $v_{p,s,min}^{v_{X,s}=0}$. For notational convenience we define $g_{p,m}^0 := g_{max,m}^{v_{p,s,min}=0} = v_{X,s,max,m}^{v_{p,s,min}=0}$ and $v_{p,s,m}^0 := v_{p,s,min,m}^{g_m=0} = v_{p,s,min,m}^{v_{X,s,m}=0}$. In the current work, we modified the definition²¹ so that it is always positive by defining it as the following:

$$GCS_{p,s,m} = \begin{cases} \frac{IA_m}{TA_m} \cdot \frac{Y_{P/S,min,m}^{g_{max}}}{Y_{P/S,max}} + 2 & \text{if } v_{p,s,m}^0 > 0, g_{p,m}^0 = 0 & \text{SUCP} \\ \frac{IA_m}{TA_m} \cdot \frac{Y_{P/S,min,m}^{g_{max}}}{Y_{P/S,max}} + 1 & \text{if } v_{p,s,m}^0 = g_{p,m}^0 = 0 & \text{dGCP} \\ \frac{IA_m}{TA_m} \cdot \frac{Y_{P/S,min,m}^{g_{max}}}{Y_{P/S,max}} + 0 & \text{if } g_{p,m}^0 > 0 & \text{wGCP} \end{cases}$$

which always has a positive value. Because the model has a positive nongrowth-associated ATP maintenance (NGAM) value, we do not have to distinguish cases when SUCP could start at the origin. Higher

GCS values within each tier are better as more of the total area is excluded and/or yields are closer to the theoretical maximum. We used numerical integration to compute the areas under the production envelope curves. Tables with all metrics for all RE design solutions are provided in Supporting Information, Data S3. The Jaccard ratio was computed as $\frac{A \cap B}{A \cup B}$ for the two sets A and B being compared.

2.6 | Computational implementation

Computations for this research were performed on the Pennsylvania State University's Institute for Computational and Data Sciences' Roar supercomputer. FBA calculations, Flux Coupling Finder, and OptKnock were implemented using GAMS 33.2.0 using IBM ILOG CPLEX solver. Modeling strain design, adding non-native reactions, set definitions, production envelopes and data processing were computed using the COBRApy package (version 0.13.4),⁴⁴ pandas,⁴⁵ and used IBM ILOG CPLEX solver (version 12.9). Visualization of hypergraphs used HyperNetX.⁴⁶

3 | RESULTS AND DISCUSSION

3.1 | Selection of substrates and products

Although *I. orientalis* is capable of growth on a range of carbon substrates, for our analysis we focused on the two major sugar monomers derived from components in processed hydrolyzed biomass^{47,48}—glucose and xylose—which can be >90% of it by weight.⁴⁹ As noted earlier, wild-type *I. orientalis* grows well on glucose, and grows on xylose after a delay.²⁶ Thus, we selected both glucose and xylose as the carbon substrates as proxy constituents of hydrolyzed biomass in the study.

Over a decade ago, the US Department of Energy identified twelve building block chemicals that can be produced from sugars⁵⁰ and these platform chemicals remain relevant to date. Eight of these building block chemicals are organic acids which range in length from three to six carbons. They are 3-hydroxypropionic (i.e., 3-hydroxypropanoic), fumaric, malic, succinic, itaconic, levulinic (i.e., 4-oxopentanoic), 2,5-furandicarboxylic and muconic acids. This report delineated their subsequent conversion to high-value bio-based chemicals and materials. For example, 3-hydroxypropionic acid can be converted into chemicals such as 1,3-propanediol, acrylic acid, methyl acrylate, and acrylamide, whereas succinic acid (i.e., butanedioic acid) can readily be converted to polymer precursors such as 1,4-butanediol, N-methyl-2-pyrrolidone, tetrahydrofuran and γ -butyrolactone. In addition to these building block chemicals, a number of other small organic acids have important uses and markets. For example, lactic acid is used as a monomer for polymers,²⁸ 3-hydroxy-3-methylbutanoic acid (HMB) is used as a human dietary supplement,⁵¹ and 2,4-dihydroxybutyric acid is used for chemical synthesis of the methionine analogue 2-hydroxy-4-(methylthio)butyrate used in animal feed.⁵²

The products examined in the current study were largely drawn from the study from the US Department of Energy⁵⁰; in addition to

TABLE 1 List of organic acids examined in the current study and computed maximum theoretical yields with respect to carbon balances on the indicated substrate

Name	ID ^a	Formula	Charge	Glucose			Xylose		
				Y_{P_c/S_c}	Y_{P_e/S_e}	$Y_{P/S}$	Y_{P_c/S_c}	Y_{P_e/S_e}	$Y_{P/S}$
Malonate	malon	C ₃ H ₂ O ₄	-2	2	3	2	1.667	2.5	1.667
Pyruvate	pyr	C ₃ H ₃ O ₃	-1	2	2.4	2	1.667	2.0	1.667
D-Lactate	lac__D	C ₃ H ₅ O ₃	-1	2	2	1.876	1.667	1.667	0.520
3-Hydroxypropanoate	3hpp	C ₃ H ₅ O ₃	-1	2	2	1.994	1.667	1.667	0.553
Fumarate	fum	C ₄ H ₂ O ₄	-2	1.5	2	1.5	1.25	1.667	1.25
L-Malate	mal__L	C ₄ H ₄ O ₅	-2	1.5	2	1.5	1.25	1.667	1.25
Succinate	succ	C ₄ H ₄ O ₄	-2	1.5	1.714	1	1.25	1.428	0.833
2-Oxobutanoate	2obut	C ₄ H ₅ O ₃	-1	1.5	1.5	1.367	1.25	1.25	1.13
2,4-Dihydroxybutanoate	24dhbut	C ₄ H ₇ O ₄	-1	1.5	1.5	1.428	1.25	1.25	1.146
3-Hydroxybutanoate	bhb	C ₄ H ₇ O ₃	-1	1.5	1.333	1	1.25	1.111	0.827
4-Hydroxybutanoate	ghb	C ₄ H ₇ O ₃	-1	1.5	1.333	1	1.25	1.111	0.833
2-Oxoglutarate	akg	C ₅ H ₄ O ₅	-2	1.2	1.5	1	1	1.25	0.833
Itaconate	itacon	C ₅ H ₄ O ₄	-2	1.2	1.333	1	1	1.111	0.833
Citramalate	citm	C ₅ H ₆ O ₅	-2	1.2	1.333	1	1	1.111	0.833
D-Xylonate	dxylnt	C ₅ H ₉ O ₆	-1	1.2	1.333	0	1	1.111	1
2-Oxopentanoate	2oxptn	C ₅ H ₇ O ₃	-1	1.2	1.091	1	1	0.909	0.827
4-Oxopentanoate	4oxptn	C ₅ H ₇ O ₃	-1	1.2	1.091	1	1	0.909	0.818
3-Methyl-2-oxobutanoate	3mob	C ₅ H ₇ O ₃	-1	1.2	1.091	1	1	0.909	0.810
3-Hydroxy-3-methylbutanoate	3hivac	C ₅ H ₉ O ₃	-1	1.2	1	0.667	1	0.833	0.556
3-Hydroxypentanoate	3hpt	C ₅ H ₉ O ₃	-1	1.2	1	0.891	1	0.833	0.706
Citrate	cit	C ₆ H ₅ O ₇	-3	1	1.333	1	0.833	1.111	0.833
Muconate	ccmuac	C ₆ H ₄ O ₄	-4	1	1.091	0.857	0.833	0.909	0.675

^aID taken from BiGG Models when available.

the eight organic acids given above, we expanded the list to include the organic acids present in the report's down selection of the top 30 results, excluding 2,5-furandicarboxylic acid. We also included a number of other small organic acids that have been examined for production in *E. coli*,^{20,53,54} or have important uses and markets. The final list of 22 target products is given in Table 1, arranged by increasing number of carbons, and the location of these products is shown in context of the metabolic network in Figure 2, which indicates those requiring heterologous enzymes be added to the model as summarized in Table 2. When multiple potential pathways existed, we chose those with the highest number of identified and characterized enzymes.

3.2 | Production potential

When evaluating the potential of forming a product from a substrate, factors from both chemistry and biochemistry weigh in. We performed an initial analysis of the three theoretical yields (two chemical and one biochemical) from conversion of the substrates into each product. The first was carbon balance yield, Y_{P_c/S_c} , that uses is the number of carbons in the product and sugar carbon substrate. Second

was the yield based on available electrons, Y_{P_e/S_e} , which considers the charges of the products and substrate and computes the amount of oxygen required to balance an oxidation reaction of each product in its charged form. Third was the biological theoretical yield, $Y_{P/S}$, which requires placing production in context of the metabolic network of an organism. Here, we used the recent GSM model *ilsor850* that accounts for the metabolic capabilities of *I. orientalis* SD108; the model has a customized biomass reaction determined from experimental data and is mass and charge balanced.³¹ We set the oxygen uptake rate upper limit to the minimal oxygen uptake that does not impinge the maximum growth rate of the wild-type model, and we note that *I. orientalis* is incapable of anaerobic growth.²⁵

3.3 | Chemical theoretical yields

We first performed a carbon balance, denoted as Y_{P_c/S_c} in Table 1, which was simply the ratio of carbons in the C3–C6 organic acid products and glucose (C6) or xylose (C5). We used these values to order the products in the table and as a metric for comparing the other yield computations. As seen in Table 1, for the yields computed from available electrons, Y_{P_e/S_e} , can differ for organic acids with the same

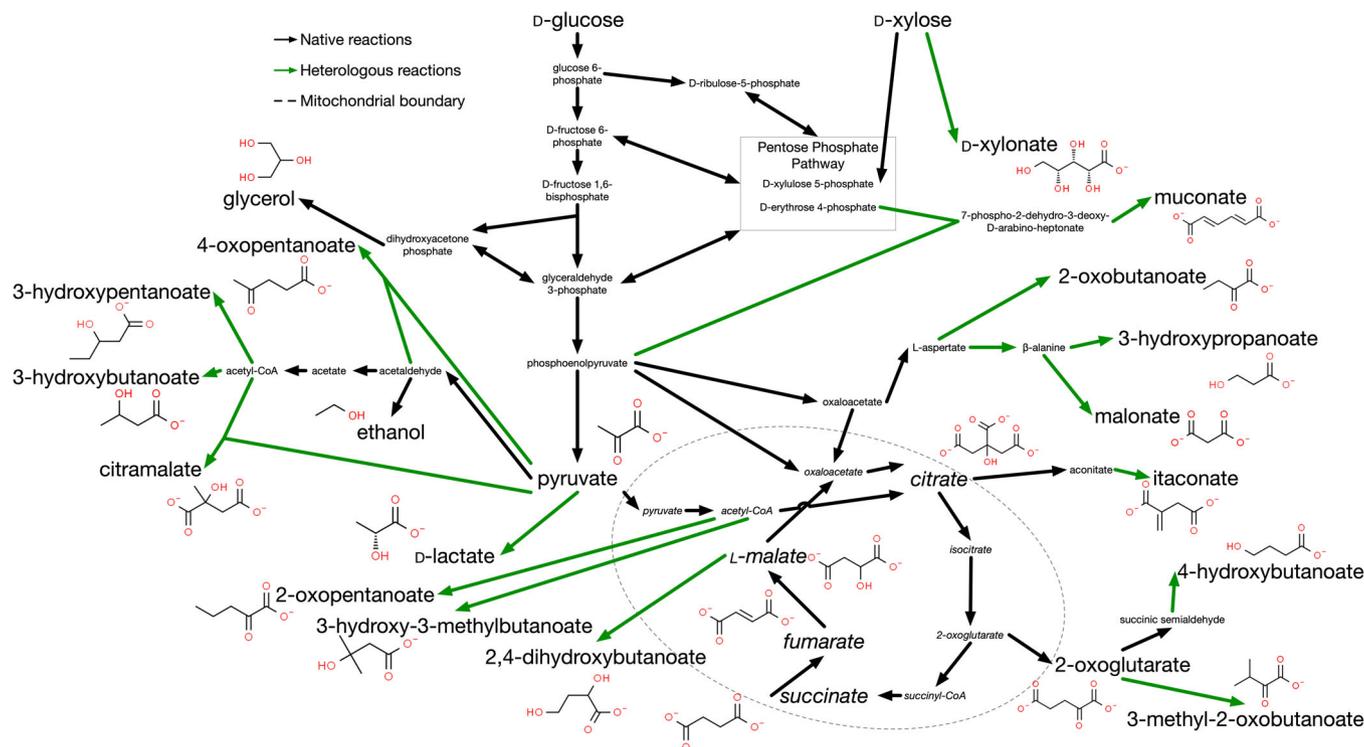


FIGURE 2 Connections of target organic acids to *I. orientalis* metabolism. Pathways involving heterologous reactions added to the network are indicated in green. Arrows indicate overall reaction pathway reversibility, with multiple reactions condensed into a single line. The targeted organic acids are displayed with their chemical structures, as well as two possible liquid by-products, glycerol, and ethanol.

number of carbons and generally decreases as the number of carbons in the product increases, reflecting the net charge on these mostly mono or dicarboxylic acids. By depending on the structure and charge of each product, these chemical theoretical yields form the upper bound of what is achievable in the absence of additional reducing power. Because they permit electron-balanced carbon uptake from CO_2 in the form of HCO_3^- as a reactant in the equation, Y_{P_e/S_e} can be greater than Y_{P_c/S_c} since this yield is from a sugar substrate carbon standpoint. Notably, the 11 products with $Y_{P_e/S_e} > Y_{P_c/S_c}$ highlight potential opportunities for engineering carbon fixing strategies. For seven of the products Y_{P_e/S_e} is less than Y_{P_c/S_c} , reflecting a lower potential, at least without a source of additional reducing power. For these cases, Y_{P_e/S_e} instead of Y_{P_c/S_c} tempers expectations of what could be theoretically achievable for any biological system, without other sources of reducing power.

3.4 | Biochemical theoretical yields

Examination of the model revealed that six target metabolites could be produced by *I. orientalis* without modifications to the model and include products such as succinic acid, pyruvic acid, L-malic acid, and citric acid. For the remaining product targets, we separately implemented pathways that enable their production by adding mass and charge balanced reactions to the network. The connection of the products to metabolism are highlighted in Figure 2 and the added

pathways are outlined in Table 2. For instance, for itaconic acid we added cis-aconitic acid decarboxylase (CAD) and associated transporters and exchange fluxes. For 3-hydroxypropionic acid we added the non-native reactions in the β -alanine pathway (i.e., pathway III⁵⁵). The full specifics of all these pathways are given in Supporting Information, Data S1. At this stage, no other modifications were considered such as reaction eliminations or additions to other parts of the network beyond that minimally required to enable metabolite production, and we performed each computation without any additional constraints other than the model's value for nongrowth-associated ATP maintenance. The product D-xylonate was not able to be produced from glucose and was only examined using xylose substrate conditions.

Using models with added reactions, as required for product formation, we used flux balance analysis (FBA)¹⁴ to compute the maximum biological product yields, $Y_{P/S}$. As seen in Table 1 for glucose six acids can be produced at the maximum carbon balance yield and for xylose, seven can. Four of the targeted organic acids could be produced at biochemical yields close to the carbon balances for both sugars: 3-hydroxypropanoate, 2,4-dihydroxybutanoate, D-lactate, and 2-oxobutanoate. Interestingly, these same four are the ones for which the corresponding carbon balance and available electron balance have the same value. The lower yield from the biological analysis is not unexpected, at least in part, for some products because the lower yield stems from the non-growth associated ATP maintenance that diverts a small amount of carbon through ATP production irrespective

TABLE 2 Heterologous reactions added to enable the production of the indicated organic acid product

Product	Added reactions ^a
Malonate	Aspartate 1-decarboxylase (ASP1DC); beta-alanine-pyruvate aminotransferase (ALABAT); malonic semialdehyde oxidoreductase (MSADx, MSADy)
D-Lactate	Lactate dehydrogenase (LDH_D)
3-Hydroxypropanoate	Aspartate 1-decarboxylase (ASP1DC); beta-alanine-pyruvate aminotransferase (ALABAT); malonic semialdehyde reductase (MSAR)
2-Oxobutanoate	Transporter
2,4-Dihydroxybutanoate	Malate kinase (MALK); Malate semialdehyde dehydrogenase (MASD); malate semialdehyde reductase (MALSARx, MALSARy)
3-Hydroxybutanoate	Acetoacetyl-CoA hydrolase (AACOAH); 3-hydroxybutanoate oxidoreductase (BDH)
4-Hydroxybutanoate	Gamma-hydroxybutyrate dehydrogenase (GHBDHx, GHBDHy)
Itaconate	Cis-aconitic acid decarboxylase (CAD)
D-Xylonate	D-Xylose NADP+ 1-oxidoreductase (XYLOR); xylonolactonase (XYLC)
2-Oxopentanoate	3-Hydroxyacyl-CoA dehydrogenase (acetoacetyl-CoA) (HACD1); 3-hydroxyacyl-CoA dehydratase (3-hydroxybutanoyl-CoA) (ECOAH1); acyl-CoA dehydrogenase (butanoyl-CoA) (ACOAD1f); 2-oxopentanoic acid decarboxylase (2OXPTNDH)
4-Oxopentanoate	4-Hydroxy-2-oxopentanoate aldolase (HOPNTAL); 4-hydroxy-2-oxo-pentanoate reductase (R4H2OPNTNR); 2,4-dihydroxy-pentanoate dehydratase (24DHPNTADH and 4O2PNTNR)
3-Methyl-2-oxobutanoate	Transporter
3-Hydroxy-3-methylbutanoate	4-Hydroxyphenylpyruvate dioxygenase/alpha-ketoisocaproate dioxygenase (RE1266C)
3-Hydroxypentanoate	Acetyl-CoA:2-oxobutanoate C-acetyltransferase (AC2OBUTAT); 2-ethyl-2-hydroxybutanedioate carboxy-lyase (2E2HOBTAEBOX)
Muconate	3-Dehydroshikimate dehydratase (DHSKDH); protocatechuic acid (PCA) decarboxylase (PCADC); catechol dioxygenase (CATADOX)

^aTransporters are only listed if the sole addition. Exchange fluxes are not listed. Complete details of added pathways are in Supporting Information, Data S1.

of other processes in the model. The remaining products had lower yields, with 3-hydroxy-3-methylbutanoate having a substantially low one. Via FBA, we found that all of the selected organic acids have the potential under some condition to be homofermenting, with the only byproduct being CO₂. We also found that, as expected, enforcing a minimal specific growth rate, $v_{X,S}$, of 10% that of the wild-type maximal specific growth rate corresponding to a carbon source uptake of uptake of 10 mmol gDW⁻¹ h⁻¹ (i.e., growth rate of at least 0.1033 h⁻¹ for growth on glucose or 0.0855 h⁻¹ for growth on xylose) resulted in correspondingly decreased yields.

3.5 | Potential from augmentation

Noting that many yields could use improvement, we examined the impact of additions to the network that could increase yield. Specifically, for each organic acid we examined if the addition of fumarate reductase (FRD) activity would improve yield, which has been used experimentally²⁷ and in silico³¹ to increase succinic acid production in *I. orientalis*. We also examined if the uptake of carbonate, HCO₃⁻, facilitated by pyruvate carboxylase (PC) activity which has increased citric acid production in *Yarrowia lipolytica*.⁵⁶ We found that only succinic

acid production was improved by fumarate reductase activity, and only malate and fumarate were improved by carbonate uptake, as was succinic acid but only with simultaneous expression of fumarate reductase activity. Notably, both malate and fumarate were able to reach the available electron theoretical yields for glucose, a 33% increase, as was succinate for both glucose and xylose via simultaneous carbonate uptake and fumarate reductase expression, a 71% increase.

3.6 | Oxygen dependence

We generated production envelopes for each target product to examine how restricting oxygen uptake rates impacts product exchange flux rates. We illustrate these results in Figure 3 for glucose and xylose using the wild-type or augmented networks without any imposed reaction eliminations. We varied the oxygen uptake from 0 (i.e., anaerobic, under which conditions *I. orientalis* does not grow) to 18.18 mmol gDW⁻¹ h⁻¹ (i.e., the minimal oxygen uptake that does not impact the maximum growth rate on a 10 mmol glucose gDW⁻¹ h⁻¹ basis). In general, for the nonaugmented pathways, the products for which Y_{P_e/S_e} is less than Y_{P_e/S_e} required higher oxygen

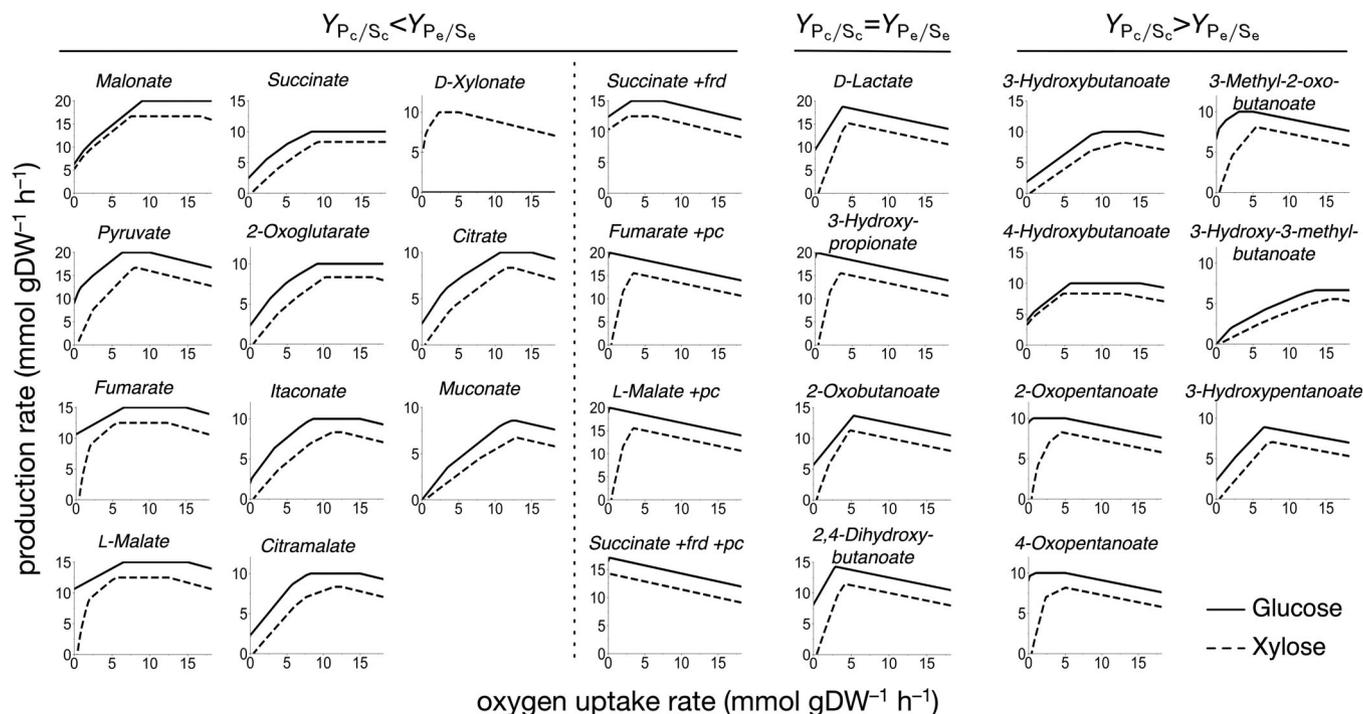


FIGURE 3 Oxygen utilization using glucose (solid line) and using xylose (dashed line) as the substrate. The products are arranged by number of carbons and by a comparison of carbon and available electron yields. A dotted line separates production envelopes for the organic acids augmented with FRD and PC expression.

uptake rates to achieve the maximum production rate, whereas those for which Y_{P_c/S_c} equals Y_{P_e/S_e} had relatively low oxygen uptake requirements (i.e., below $5 \text{ mmol gDW}^{-1} \text{ h}^{-1}$). In particular, 3-hydroxypropionic acid had its maximum value at near anaerobic conditions. The impact on oxygen requirements is improved for all the augmented pathways when compared to the native pathways, which is especially pronounced for the uptake of HCO_3^- via PC. For xylose, the results are similar but typically shifted to higher oxygen requirements. Of particular note is that some products, including 3-hydroxypropionic acid, using xylose as the carbon substrate have an oxygen threshold below which they cannot be produced. This situation arises from inability of *I. orientalis* to grow anaerobically and the resulting incapacity to balance cofactors under those conditions. We also find that the oxygen utilization for succinate production has low oxygen requirements in the context of FRD and PC expression.

3.7 | Network analysis and accessible reactions for targeted eliminations

At the onset of our analysis, we performed analyses to reduce the reaction space for subsequent analyses to those that are accessible and practical as targets for genetic knockout. The first stage was performed in common to all carbon source inputs, whereby examination of the model's 1832 reactions found 580 have Systems Biology Ontology (SBO)⁴² terms for transporters and 173 have SBO terms for exchange or other boundary reactions. As many as 719 have no

defined gene-protein-reaction (GPR) associations or are spontaneous, which can overlap with the previous sets. For aerobic glucose conditions, by using FVA we found 760 blocked reactions and 386 essential reactions. Similarly, for aerobic xylose conditions we found 755 blocked and 391 essential reactions. Because we seek to find implementable interventions, we then examined the set of essential genes and the reactions they encode; some of these reactions might not be essential individually, but knocking out a gene could potentially remove several reactions at once thereby forming a synthetic lethal. Others are not essential but are only associated with a gene that is essential because it catalyzes a different reaction which is essential. We identified 230 essential genes for glucose and 239 for xylose, which led to 98 (glucose) and 97 reactions (xylose) non-essential reactions encoded by only an essential gene. Combining these results with a subsequent examination of reactions with complex GPR identified a total of 106 (glucose) and 107 (xylose) reactions excluded by the GPR analysis. By combining the results for classification, blocked, and essential analyses, we identified a set of potentially reactions accessible to be eliminated containing 279 (glucose) and 280 (xylose) reactions.

Optimal reaction elimination algorithms such as OptKnock are computationally expensive and memory intensive as the number of variables and allowed simultaneous eliminations increase. To further improve computational tractability by reducing the number of reactions targeted for elimination, we turned to flux coupling analysis by using the Flux Coupling Finder (FCF) algorithm.⁴⁰ Doing so allows us to identify equivalent knockouts, and reduce the reaction space for

subsequent analyses. We found 81 (glucose) and 82 (xylose) sets of fully coupled reactions. By permitting only one member of each to represent the group, taking into account the previous excluded reactions, we found a set of non-redundant practical reactions for glucose and xylose to both be 223 reactions, which differ by six reactions: ferrocyclochrome-c:oxygen oxidoreductase (FECOOR_m), ubiquinol:ferricytochrome c reductase (FEQRq7_m) and glucose-6-phosphate isomerase (PGI) present for glucose and aldehyde dehydrogenase (3-aminopropanal, NAD) (ALDD22x), hexokinase (D-glucose:ATP) (HEX1) and spermine synthase (SPRMS) present for xylose; these six are essential reactions for the other substrate. These sets are made available in Supporting Information, Data S2.

For each substrate-product pair, we used Flux Variability Analysis (FVA) to identify sets of reactions in the non-redundant practical reactions that individually permitted zero flux through each when constraining the production of the target to its maximum value obtained when biomass production was constrained to at least 10% of its maximum value, that is, $v_{j,\min}^{g=0.1g_{\max}} \leq 0 \leq v_{j,\max}^{g=0.1g_{\max}}$. For some, zero was the only permitted value (i.e., any nonzero flux through them would negatively impact product yield), whereas others had non-zero $v_{j,\min}^{g=0.1g_{\max}}$ and/or $v_{j,\max}^{g=0.1g_{\max}}$. We also used FBA to examine the effect on product yield from eliminating each reaction in the respective non-redundant practical reactions. Most had no impact and the vast majority of those that did only did so with minimal reductions (i.e., < 1%). Some, such as pyruvate decarboxylase (PYRDC), glucose-6-phosphate isomerase (PGI) and NADH:ubiquinone oxidoreductase (complex I) (NADHcplxI_c_m) had larger negative impacts (i.e., 20% or more reduction) on production for some but not all substrate-product pairs. These sets (and corresponding reductions in yield, as appropriate) are made available in Supporting Information, Data S2.

3.8 | Initial computationally predicted strain designs

We used OptKnock to determine RE designs for the target organic acids in order to broadly examine the full range of growth-coupled production strategies available, including potentially growth-coupled production. For initial OptKnock simulations, all reactions other than the 223 non-redundant practical reactions for glucose and xylose were fixed to be on and not allowed to be eliminated. We also excluded reactions for which single elimination resulted in zero product formation. All RE designs were subsequently partitioned into pGCP, wGCP, dGCP, and SUCP categories for each substrate-product pair. Excluding those for the augmented pathways, across all glucose, reaction pairs we found as many as 3674 RE solutions containing of up to five reactions that had a potential product yield of at least 10% carbon yield at maximum growth rate (i.e., $Y_{P/S,\max}^{g_{\max}}/Y_{P_c/S_c} \geq 0.1$); these RE solutions contained 105 different reactions. For xylose we found as many as 3318 solutions involving 95 reactions meeting the same criterion. These reactions are provided in Supporting Information, Data S3. Of these, 13 (glucose) and 3 (xylose) reactions were unique to the one substrate. The number of designs per substrate-product

pair varied considerably, with some such as native metabolites pyruvate, fumarate, and succinate having many solutions. For glucose, we found dGCP or SUCP designs for only six of the products (viz., malonate, pyruvate, D-lactate, 2-oxoglutarate, and 2-oxopentanoate) and for xylose, we found dGCP or SUCP designs for malonate, D-lactate, succinate, 2-oxoglutarate, D-xylonate, and 2-oxopentanoate, with succinate, 2-oxoglutarate, and D-xylonate having numerous solutions. Interestingly, D-xylonate had the largest number of SUCP designs, reaching as high as 0.923 of the carbon yield from xylose. We found wGCP designs for 15 and 13 products from glucose and xylose, respectively. Only pGCP designs existed for malate, 4-hydroxybutanoate, and 4-oxopentanoate for both sugars, and as many as 10 substrate-product pairs had no solutions, including 2-oxobutanoate, 3-hydroxybutanoate, citramalate, 3-hydroxypentanoate. As seen in Figure 2, the later three all derive wholly or in part from acetyl-CoA. For each strain design we computed our evaluation metrics including $Y_{p,\min,S}^{g_{\max,S}}$, $Y_{p,\max,S}^{g_{\max,S}}$, g^0 , $Y_{p,\min,S}^{g=0}$, $SSP_{p,S}$, and $GCS_{p,S}$, as well as the range of oxygen uptake at $g_{\max,S}$. We ranked the RE solutions descending using $Y_{p,\max}^{g_{\max}}$ for pGCP and $Y_{p,\min}^{g_{\max}}$ for wGCP, dGCP, and SUCP; the RE solutions and corresponding metrics values are provided in Supporting Information, Data S3. We compared the similarity of solutions for the two carbon sources by computing the Jaccard index across all products, shown in Table 3. Comparisons of solutions made within a substrate are in Supporting Information, Data S4. In general, we found that there was less than 50% overlap of designs for the two sugars.

3.9 | Examination of factors contributing to designs

We examined the reactions that occurred within the solutions of each product for the substrate-product pairs by creating incident matrices encompassing the reactions for each qualitative category of solutions. We found that some reactions were ubiquitous for a given substrate-product pair, or nearly so. As given in Table 4, we found that for as many as 13 of the products there were such reactions. Some of these 28 reactions, such as ATP synthase (APTS_m), pyruvate decarboxylase (PYRDC), appeared in more than one product. We also observed differences between the two substrates which was particularly noticeable for 3-hydroxypropionate, for which had mutually exclusive lists, with ribulose 5-phosphate 3-epimerase (RPE) only occurring for xylose and aspartate transaminase (ASPTA), APTS_m, and NADH dehydrogenase (NADHq7) only occurring for glucose.

Analysis for reactions that we found occurring ubiquitously in the solutions revealed enzymes that have been the target of industrial-based processes or previous strain designs. For instance, (R)-lactate:ferricytochrome-c 2-oxidoreductase has been reported as a target for improving lactic acid production in recombinant yeast.²⁹ In *E. coli*, aspartate transaminase activity has been a target for deletion to increasing the production of β -alanine,⁵⁷ which is an intermediate in the 3-hydroxypropionate pathway we used. A strain of *S. cerevisiae* that oxidatively produces succinate involved the deletion of succinate dehydrogenase.⁵⁸ The deletion of pyruvate decarboxylase, as seen in

TABLE 3 Jaccard similarity coefficient for reaction elimination solutions using glucose and xylose as substrates

Glc	Xyl															
	malon	pyr	lac_D	3hpp	fum	mal_L	succ	ghb	akg	2oxptn	4oxptn	3mob	3hivac	cit	ccmuac	dxylnt
malon	0.451	0.010	0	0	0	0	0	0	0	0	0	0.025	0	0	0.003	0.018
pyr	0.020	0.444	0	0	0	0	0	0	0	0.219	0.337	0	0.017	0	0.001	0.004
lac_D	0	0	0.421	0	0	0	0	0	0	0	0	0	0	0	0	0
fum	0	0	0	0	0.194	0	0.007	0	0	0	0	0	0	0	0.003	0.011
mal_L	0	0	0	0	0.004	0.333	0	0	0	0	0	0	0	0	0	0
succ	0	0	0	0	0.008	0	0.268	0	0.002	0	0	0	0	0	0	0.005
ghb	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
akg	0	0	0	0.022	0	0	0	0.206	0.208	0	0	0	0	0	0	0.009
2oxptn	0	0.217	0	0	0	0	0	0	0	0.372	0.219	0	0.023	0	0.011	0.019
4oxptn	0	0.314	0	0	0	0	0	0	0	0.170	0.352	0	0.014	0	0	0
3mob	0.036	0	0	0	0	0	0	0	0	0	0	0.204	0	0	0	0.010
3hivac	0.011	0.026	0	0	0	0	0	0	0	0.026	0.019	0	0.237	0	0.008	0.021
cit	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
ccmuac	0.004	0	0	0	0	0	0.001	0	0	0.006	0	0	0.002	0	0.249	0.014

our malate and pyruvate designs, has also been examined in *S. cerevisiae*,⁵⁹ although the authors focused on its impact in succinic acid production; it does occur in our designs, albeit not ubiquitously. The deletion of ATP synthase has been observed to be a frequent deletion target in other growth-coupling strategies in *E. coli*²¹ whereby the ATP yield from the pathway producing the target product is greater than the other residual ATP forming pathways after ATP synthase is deleted. In general, these ubiquitous and near-ubiquitous reactions are early candidates for implementation in *I. orientalis*, especially those that occur for several organic acids, such as pyruvate decarboxylase. As expected from our previous limited study on adding fumarate reductase (FRD) activity to improve succinic acid yield in silico,³¹ which agreed with in vivo experiments²⁷ and in the current work, we found that its production could be enhanced through the uptake of carbonate. The RE designs for succinate, fumarate and malate augmented with FRD and PC expression were similar to those without.

Using the incident matrices, the structure of RE design solutions for a substrate-product pair can be revealed, for instance, using hypergraphs⁴⁶ and bipartite graphs with nodes representing the reactions in a solution as well as a node for each solution. In general, we found that higher order sets are not necessarily constructed by simply adding on reactions to existing solutions, though such smaller sets can be kernels for some of them. For example, for sGCP designs for xylose-succinate, the five solutions of size three could be extended into 23 of the 40 solutions of size four upon the addition of one differing extra reaction. Similarly, pGCP designs can be promoted to wGCP or sGCP through the addition of reactions in higher-order sets, as can wGCP to dGCP or SUCP designs.

3.10 | Improving strain designs categorization

Using the results from Table 4 and the sets of reactions found in Supporting Information, Data S2 that impact each product's yield and from the initial computationally predicted strain designs, we subsequently examined higher-order simultaneous reaction eliminations focusing on finding designs with improved qualitative and/or quantitative metrics: either entry into a better category of GCP for those with at best \emptyset GCP, pGCP, or wGCP designs or improved quantitative metrics for those already with dGCP or SUCP designs. We also continued searches without any additional constraints on selectable reactions. Table 5 summarizes our results for all strain design solutions. We found as many as 3841 solutions involving 106 reactions for glucose and 3393 solutions and 95 reactions for xylose. These reactions are tabulated in Supporting Information, Data S2.

We found that the constrained maximum oxygen bound described above in the oxygen dependence analysis (i.e., 18.18 mmol gDW⁻¹ h⁻¹) was reached by 2498 glucose-product pair designs, of which 1037 also had the minimum oxygen uptake at the same limit. In concordance with the oxygen dependence analysis, examination revealed that except for succinate, nearly all sGCP designs had maximum oxygen uptake rates below this constraint. Results for xylose were

TABLE 4 Ubiquitous and near-ubiquitous reactions occurring in GCP solutions

ID ^a	Glucose		Xylose	
	Total #	Ubiquitous ^b	Total #	Ubiquitous
All GCP				
malon	311	PYRDC	194	PYRDC (186)
pyr	531		520	ATPS_m (508)
lac_D	79	RLFC2O, RLFC2O_m, TPI	56	RLFC2O, RLFC2O_m, TPI
3hpp	10	ASPTA, ATPS_m, NADHq7	68	
fum	427	FUM	231	RPE (67)
mal_L	2	CRNOAT, MDH	6	FUM (229)
succ	787		727	G6PDH2i (5)
24dhbut	10	GLUDy, MDH, NADHcplxl_m	0	SUCDq7_m (676)
ghb	14	ASPTAI_m, NADHq7, SUCDq7_m	14	
akg	298		271	ASPTAI_m, NADHq7, SUCDq7_m
dxyInt	N/A		703	GLUDC
4oxptn	630	ATPS_m	410	G6PDH2i (235)
3mob	338		145	G6PDH2i (690)
wGCP+sGCP				PDH_m (137)
malon	306	PYRDC	181	PYRDC
pyr	62	PYRDC	50	PYRDC
lac_D	63	RLFC2O, RLFC2O_m, TPI	47	RLFC2O, RLFC2O_m, TPI
3hpp	10	ASPTA, ATPS_m, NADHq7	0	
fum	427	FUM	229	FUM
succ	779		681	SUCDq7_m (677)
24dhbut	10	GLUDy, MDH, NADHcplxl_m	0	
akg	290	GLUDC	249	GLUDC
dxyInt	N/A		696	G6PDH2i (687)
2oxptn	0		276	NADHcplxl_m (271)
3mob	4	3MOBDC, MDH_m, NADHq7, PDH_m	0	
3hvac	38	4MOPDC, NADHcplxl_m	14	4MOPDC, NADHcplxl_m

^aID taken from BiGG Models when available.^bReaction abbreviations are: 3MOBDC 3-methyl-2-oxobutanoate decarboxylase, 4MOPDC 4-methyl-2-oxopentanoate decarboxylase, ASPTA aspartate transaminase, ASPTAI_m aspartate transaminase, ATPS_m ATP synthase, CRNOAT carnitine O-acetyltransferase, DDPA 3-deoxy-D-arabino-heptulosonate 7-phosphate synthetase, FUM fumarase, G6PDH2i glucose 6-phosphate dehydrogenase, GLUDC glutamate decarboxylase, GLUDy glutamate dehydrogenase (NADP), ICdHx_m isocitrate dehydrogenase (NAD+), ICdHyi isocitrate dehydrogenase (NADP), ICL_1 isocitrate lyase, MDH malate dehydrogenase, MDH_m malate dehydrogenase, ME1_m malic enzyme (NAD), ME2_m malic enzyme (NADP), NADHcplxl_c_m NADH:ubiquinone oxidoreductase (complex I), NADHq7 NADH dehydrogenase, PDH_m pyruvate dehydrogenase, PGI glucose-6-phosphate isomerase, PYRDC pyruvate decarboxylase, RLFC2O (R)-lactate:ferricytochrome-c 2-oxidoreductase, RLFC2O_m (R)-lactate:ferricytochrome-c 2-oxidoreductase, RPE ribulose 5-phosphate 3-epimerase, SUCDq7_m succinate dehydrogenase (ubiquinone-7), TPI triose-phosphate isomerase. Reaction abbreviations ending with _m signify reactions localized to the mitochondria.^cNumbers in parentheses are the incidence counts of the respective reaction.

TABLE 5 Classification and quantification of strain design solutions

ID ^a	Glucose															
	pGCP				wGCP				dGCP				SUCP			
	#	$Y_{P/S,max}^{f,max}$	#	$Y_{P/S,max}^{f,max}$	#	$Y_{P/S,max}^{f,max}$	#	$Y_{P/S,max}^{f,max}$	#	$Y_{P/S,max}^{f,max}$	#	$Y_{P/S,max}^{f,max}$	#	$Y_{P/S,max}^{f,max}$		
malon	5	1.729	288	1.743	0	1.809	13	1.443	165	1.268	0	1.268	0	1.450		
pyr	469	1.842	61	1.803	0	1.822	470	1.325	50	1.299	0	1.299	0	0		
lac_D	16	0.638	70	0.918	6	0.915	9	0.558	43	0.610	0	0.610	0	0.637		
3hpp	3	1.058	46	1.189	0	0	68	1.016	14	1.144	0	1.144	0	7		
fum	0		427	0.578	0	0	2	0.200	229	0.589	0	0.589	0	0		
mal_L	20	0.235	0	0	0	0	31	0.229	0	0	0	0	0	0		
succ	8	0.288	4	0.727	775	0.728	46	0.587	5	0.317	547	0.713	129	0.727		
2obut	0		0	0	0	0	15	0.927	0	0	0	0	0	0		
24dhbut	0		36	0.487	0	0	0	0	0	0	0	0	0	0		
bhb	4	0.339	0	0	0	0	0	0	0	0	0	0	0	0		
ghb	20	0.682	11	0.469	0	0	23	0.569	7	0.392	0	0.392	0	0		
akg	8	0.357	278	0.660	12	0.543	22	0.481	202	0.716	0	0.716	47	0.726		
itacon	0		11	0.259	0	0	0	0	1	0.174	0	0.174	0	0		
citm	2	0.365	0	0	0	0	0	0	0	0	0	0	0	0		
dxvInt	N/A		N/A		N/A		7	0.771	464	0.906	8	0.862	224	0.923		
2oxpnt	244	0.908	162	0.298	97	0.926	221	0.658	260	0.489	4	0.160	12	0.313		
4oxpnt	630	0.921	0	0	0	0	410	0.662	0	0	0	0	0	0		
3mob	334	0.591	4	0.443	0	0	145	0.503	0	0	0	0	0	0		
3hivac	147	0.505	38	0.505	0	0	99	0.394	14	0.391	0	0.391	0	0		
3hpt	0		0	0	0	0	0	0	0	0	0	0	0	0		
cit	1	0.133	0	0	4	0.203	0	0	0	0	3	0.169	0	0		
ccmuac	2	0.131	288	0.408	0	0	4	0.104	298	0.406	0	0.406	0	0		

^aID abbreviations are as in Table 1.

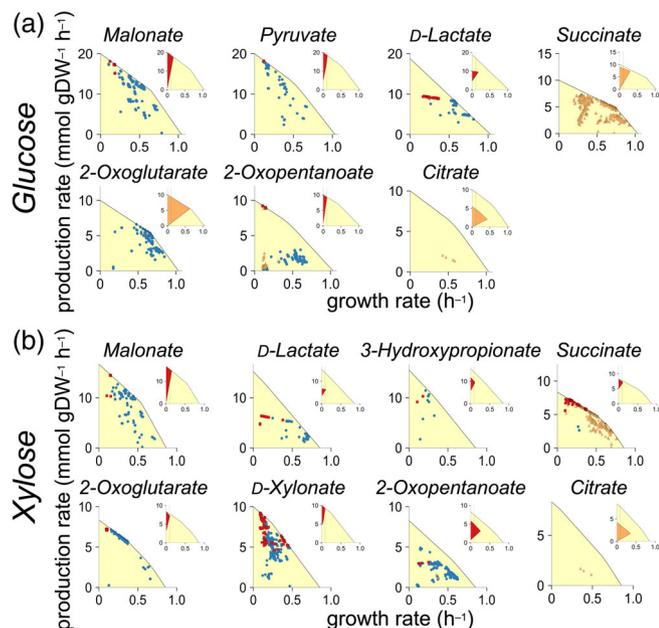


FIGURE 4 Production envelopes showing the product rates accessible for each feasible growth rate using (a) glucose and (b) xylose as the substrate. In the main of the paired plots, each design solution is represented by a scatter point at the maximum growth rate and associated minimum production rate. These designs are color coded as in Figure 1, with blue circles indicating weakly growth-coupled production designs, orange triangles directionally growth-coupled production designs, and red squares substrate-uptake coupled production designs. On the inset of the paired plots, the production envelope for the strongly growth-coupled production design with the highest GCS metric is filled with the color of its corresponding symbol.

proportionally smaller, with 1732 designs having maximum oxygen uptake rates constrained by the bound and 364 designs having the minimum oxygen uptake also so constrained. This difference arises because the oxygen uptake constraint was set based on glucose uptake; selecting xylose designs with oxygen uptake rates above $15.42 \text{ mmol gDW}^{-1} \text{ h}^{-1}$ (i.e., the minimal oxygen uptake that does not impact the maximum growth rate on a $10 \text{ mmol xylose gDW}^{-1} \text{ h}^{-1}$ basis) reveals 2695 designs (maximum) and 1368 designs (minimum).

We added 3-hydroxypropionic and citric acids to the set of acids with sGCP designs, added 4-oxobutanoic acid to those with wGCP, and identified solutions with increased quantitative metrics for other acids. We were able to find one pGCP design for citramalate from glucose and one for 2-oxobutanoate from xylose, but were unable to find any solutions for the other substrate-product pairs previously with only 0GCP. Figure 4 summarizes the wGCP and sGCP designs for the 15 substrate-product pairs for which we were able to find at least one dGCP or SUCP design. It also shows the corresponding production envelope for the highest-ranked design of each pair, as determined by GCS. A total of 11 substrate-product pairs had at least one SUCP design. In total, nine pairs are predicted to have yields near the target biological theoretical maximum with seven also having high

carbon yields. All design solutions summarized in Table 5 are provided in Supporting Information, Data S3, and production envelope scatter plots for all organic acids are provided in Supporting Information, Data S5.

For the highest-ranked RE designs for the 15 pairs in Figure 4, we used GPR associations to ascertain the feasibility of implementing them via gene knockouts. For each pair, we report the corresponding number of reaction:gene interventions; the differing numbers for each intervention type arose typically because of isozymes associated with some reactions. For glucose, five products had identical production envelopes for both RE and the corresponding gene knockouts (viz., malonate (5:5), D-lactate (5:7), succinate (5:7), 2-oxoglutarate (5:5), and 2-oxopentanoate (3:4)). For xylose, three had identical production envelopes for both RE and gene knockouts (viz., malonate (5:5), D-lactate (5:7), and 3-hydroxypropanoate (6:7)). Glucose production of pyruvate (5:6) and xylose production of 2-oxoglutarate (5:5) and D-xylonate (5:6) had the same $v_{p,\min}^{\text{max}}$ for both RE and gene knockouts and had lower $v_{p,\min}^0$ for each respective gene implementation but remained SUCP. For xylose production of succinate (5:7) the gene implementation revealed a synthetic lethal involving glycerol-3-phosphate acyltransferase (G3PAT_I). Use of an equivalent alternative RE design (i.e., one having all the same number of reactions, $v_{p,\min}^{\text{max}}$, and GCS score) corresponded to a 6 gene knockout design with the same $v_{p,\min}^{\text{max}}$ and a lower $v_{p,\min}^0$ that remained SUCP. Xylose production of 2-oxopentanoate (4:4) similarly had the same $v_{p,\min}^{\text{max}}$ for both RE and gene knockout designs but the latter's $v_{p,\min}^0$ decreased to zero which relegated it to dGCP. Finally, analysis of citrate produced from glucose (6:7) and xylose (6:7) also uncovered a synthetic lethal at the gene level, arising from the inclusion of the cytosolic reaction 3-deoxy-D-arabino-heptulosonate 7-phosphate synthetase (DDPA_c). Examination for both substrates of reaction and gene implementations excluding this reaction reveals that all designs remain dGCP but fall below the carbon yield cutoff filter. A more general analysis using FBA to uncover other synthetic lethals in other RE designs also revealed participation by the reactions glycerol-3-phosphate acyltransferase (G3PAT_rm), ribonucleotide reductase (UDP) (RNDR4_c) and 3-deoxy-D-arabino-heptulosonate 7-phosphate synthetase (DDPA_m). These five reactions impact as many as 246 glucose solutions and 314 xylose solutions. In summary, although there can be execution challenges when designing based on RE, we found realizable gene-level implementations with identical or comparable production envelopes to the RE designs for 12 of the 15 substrate-product pairs that exhibited sGCP, and the remaining three nevertheless retained sGCP.

4 | CONCLUSION

In this work, we detailed the potential of *I. orientalis* as an organic acid producer. We found that for most of the selected products OptKnock suggested designs involved high yields. As we noted, currently *I. orientalis* metabolizes xylose slowly after a long delay, and this ability to consume xylose is consistent with the presence in its genome of all

genes needed for D-xylose conversion to D-xylose-5-P via the oxido-reductase pathway of xylose reductase, xylitol dehydrogenase and xylulokinase, which can then funnel into the pentose phosphate pathway.³¹ However, metabolic engineering methods could be applied to improve xylose as an effective substrate. Genetic interventions or serial cultivation could be used to modify directly or indirectly the existing genes needed for the oxido-reductase pathway along with their regulation. We note that metabolic engineering methods have permitted the yeast *Saccharomyces cerevisiae* to use xylose via the oxido-reductase pathway by expressing genes from *Scheffersomyces stipites*,⁶⁰ which could similarly be implemented in *I. orientalis*. Alternatively, an isomerase pathway from another organism could be introduced that directly converts xylose into xylulose followed by the xylulokinase of the oxido-reductase pathway.

The oxygen uptake analysis underscores one facet of operational conditions and the impact it can have on production yields. Our results suggest that 3-hydroxypropionic acid production is well-suited to require little input of oxygen beyond that required to grow the necessary cell mass in a bioreactor. Similar outlooks were found for succinate, fumarate and malate in the presence of additional functionality added to the cell such as fumarate reductase or pyruvate carboxylase activity. We anticipate future examination to highlight other adjunct heterologous pathways for other products that improve oxygen requirements as well as other pathway augmentations that can increase the in vivo yield of organic acids closer to the carbon and/or available electron maximum theoretical yields.

We were able to find growth-coupled production designs or potentially growth-coupled designs for almost every product (i.e., 37 of the 43 substrate-product pairs). We found some form of unique GCP designs (i.e., wGCP or sGCP), for 29 substrate-product pairs, of which 9 had at least one sGCP design with yields near the theoretical biological maximum. Work remains, however, to move those with wGCP designs into dGCP or SUCP designs and to identify even pGCP design solutions for some pairs, including 2-oxobutanoate, 3-hydroxybutanoate, and 3-hydroxypentanoate. As seen in Figure 2, the latter two derive wholly from acetyl-CoA. The inability to couple an acetyl-CoA drain to growth was observed in *E. coli* network analysis,²¹ although there the problem was described as being due to the model's inability to compensate for the CoA drain, which similarly impacted succinyl-CoA. We were able to find one pGCP design for citramalate, which also derives in part from acetyl-CoA. In the absence of available wGCP or sGCP designs, it is advantageous to identify, as we have done here through the use of OptKnock, if there are strain designs or conditions under which the product could nevertheless be noncompeting with biomass production (i.e., pGCP) instead of directly antagonistic to biomass production (i.e., \emptyset GCP). Additional effort is required to determine effective production routes for those products which only had pGCP designs. One alternative approach to growth coupling is to use a method such as OptForce⁶¹ in combination with labeled substrates in order to determine optimal flux values for high yields and then engineer the system to have such fluxes, by up or down regulating specific fluxes. Such an approach is especially attractive for products without sGCP designs. We provide the caveat that

our results do not necessarily preclude higher order RE strategies from forming wGCP or sGCP designs. We furthermore note that additional analyses are required to map RE into implemented gene knock-outs and check feasibility. Such analyses include considerations such as GPR associations and identification of major isozymes.⁶² Here, we were able to elucidate effective GCP gene implementations for nearly all the top scoring RE designs for substrate-product pairs having sGCP designs. We point out that the flux coupling reduction process we used identifies equivalent reaction eliminations and has the potential to avoid reactions that would otherwise be prohibited because of synthetic lethals or other gene implementation difficulties.

The somewhat low Jaccard index between solutions for different substrates for the same product suggests that deriving a single strain capable of effectively metabolizing both substrates into product could be challenging, especially if much of the metabolic engineering work is primarily tested and focused at the onset on one of the sugars. Early examinations of strain design with an eye for one that targets both could behoove the process. Another approach could be to design and use mixed cultures, with one strain for each substrate. Organic acid production at low pH can have energetically expensive energy requirements for product export.⁶³ Typically GSM models, including the one used in the current work, have been reconstructed with charges on metabolites determined at a single neutral pH value across the model and with constant ATP maintenance values,⁶⁴ although some models have included physicochemical specification differences between pH 5 and 8.^{65,66} We anticipate that accounting for physicochemical specifications of different compartments, allowing pH-dependence of ATP maintenance, and incorporating regulatory constraints could further enhance predicting the production potential of *I. orientalis*. The process and methods we apply herein can be readily applied to other product categories.

AUTHOR CONTRIBUTIONS

Patrick F. Suthers: Conceptualization (equal); data curation (lead); formal analysis (lead); investigation (lead); methodology (lead); validation (lead); visualization (lead); writing – original draft (lead); writing – review and editing (equal). **Costas Maranas:** Conceptualization (equal); resources (lead); supervision (lead); writing – review and editing (equal).

ACKNOWLEDGMENTS

This work was funded by the DOE Center for Advanced Bioenergy and Bioproducts Innovation (US Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018420). Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the US Department of Energy. Funding also provided by the DOE Office of Science, Office of Biological and Environmental Research (Award Number DE-SC0018260) Computations for this research were performed on the Pennsylvania State University's Institute for Computational and Data Sciences' Roar supercomputer.

CONFLICT OF INTEREST

The authors declare no commercial or financial conflict of interest.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available in the Supporting Information of this article.

ORCID

Costas D. Maranas  <https://orcid.org/0000-0002-1508-1398>

REFERENCES

1. Bailey JE. Toward a science of metabolic engineering. *Science*. 1991; 252:1668-1675.
2. Gu C, Kim GB, Kim WJ, Kim HU, Lee SY. Current status and applications of genome-scale metabolic models. *Genome Biol*. 2019; 20(1):121.
3. Heirendt L, Arreckx S, Pfau T, et al. Creation and analysis of biochemical constraint-based models using the COBRA toolbox v.3.0. *Nat Protoc*. 2019;14(3):639-702.
4. Suthers PF, Foster CJ, Sarkar D, Wang L, Maranas CD. Recent advances in constraint and machine learning-based metabolic modeling by leveraging stoichiometric balances, thermodynamic feasibility and kinetic law formalisms. *Metab Eng*. 2021;63:13-33.
5. Yim H, Haselbeck R, Niu W, et al. Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. *Nat Chem Biol*. 2011;7(7):445-452.
6. Rodrigues C, Vandenbergh LPS, Woiciechowski AL, de Oliveira J, Letti LAJ, Soccol CR. Production and application of lactic acid. *Curr Dev Biotechnol Bioeng Prod Isolat Purif Ind Prod*. 2017;543-556.
7. Tian X, Chen H, Liu H, Chen J. Recent advances in lactic acid production by lactic acid bacteria. *Appl Biochem Biotechnol*. 2021;193(12):4151-4171.
8. Baeshen MN, Al-Hejin AM, Bora RS, et al. Production of biopharmaceuticals in *E. coli*: current scenario and future perspectives. *J Microbiol Biotechnol*. 2015;25(7):953-962.
9. Van Dien S. From the first drop to the first truckload: commercialization of microbial processes for renewable chemicals. *Curr Opin Biotechnol*. 2013;24(6):1061-1068.
10. Fong SS, Burgard AP, Herring CD, et al. In silico design and adaptive evolution of *Escherichia coli* for production of lactic acid. *Biotechnol Bioeng*. 2005;91(5):643-648.
11. Layton DS, Trinh CT. Engineering modular ester fermentative pathways in *Escherichia coli*. *Metab Eng*. 2014;26:77-88.
12. Burgard AP, Pharkya P, Maranas CD. OptKnock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol Bioeng*. 2003;84(6):647-657.
13. Edwards JS, Ramakrishna R, Palsson BO. Characterizing the metabolic phenotype: a phenotype phase plane analysis. *Biotechnol Bioeng*. 2002;77(1):27-36.
14. Orth JD, Thiele I, Palsson BO. What is flux balance analysis? *Nat Biotechnol*. 2010;28(3):245-248.
15. Maia P, Rocha M, Rocha I. In Silico constraint-based strain optimization methods: the quest for optimal cell factories. *Microbiol Mol Biol Rev*. 2016;80(1):45-67.
16. Tepper N, Shlomi T. Predicting metabolic engineering knockout strategies for chemical production: accounting for competing pathways. *Bioinformatics*. 2010;26(4):536-543.
17. Tervo CJ, Reed JL. FOCAL: an experimental design tool for systematizing metabolic discoveries and model development. *Genome Biol*. 2012;13(12):R116.
18. Xu Z, Zheng P, Sun J, Ma Y. ReactKnock: identifying reaction deletion strategies for microbial strain optimization based on genome-scale metabolic network. *PLoS One*. 2013;8(12):e72150.
19. Flowers D, Thompson RA, Birdwell D, Wang T, Trinh CT. SMET: systematic multiple enzyme targeting - a method to rationally design optimal strains for target chemical overproduction. *Biotechnol J*. 2013;8(5):605-618.
20. Feist AM, Zielinski DC, Orth JD, Schellenberger J, Herrgard MJ, Palsson BO. Model-driven evaluation of the production potential for growth-coupled products of *Escherichia coli*. *Metab Eng*. 2010;12(3): 173-186.
21. Alter TB, Ebert BE. Determination of growth-coupling strategies and their underlying principles. *BMC Bioinf*. 2019;20(1):447.
22. Schneider P, Mahadevan R, Klant S. Systematizing the different notions of growth-coupled product synthesis and a single framework for computing corresponding strain designs. *Biotechnol J*. 2021; 16(12):e2100236.
23. Abbott DA, Zelle RM, Pronk JT, van Maris AJ. Metabolic engineering of *Saccharomyces cerevisiae* for production of carboxylic acids: current status and challenges. *FEMS Yeast Res*. 2009;9(8):1123-1136.
24. Kurtzman CP, Robnett CJ, Basehoar-Powers E. Phylogenetic relationships among species of *Pichia*, *Issatchenkia* and *Williopsis* determined from multigene sequence analysis, and the proposal of *Barnettozyma* gen. Nov., *Lindnera* gen. Nov. and *Wickerhamomyces* gen. Nov. *FEMS Yeast Res*. 2008;8(6):939-954.
25. Okuma Y, Endo A, Iwasaki H, Ito Y, Goto S. Isolation and properties of ethanol-using yeasts with acid and ethanol tolerance. *J Ferment Technol*. 1986;64:379-382.
26. Hisamatsu M, Furubayashi T, Karita S, Mishima T, Isono N. Isolation and identification of a novel yeast fermenting ethanol under acidic conditions. *J Appl Glycosci*. 2006;53(2):111-113.
27. Xiao H, Shao Z, Jiang Y, Dole S, Zhao H. Exploiting *Issatchenkia orientalis* SD108 for succinic acid production. *Microb Cell Factories*. 2014;13:121.
28. Suominen P, Aristidou A, Penttila M, et al., Inventors. Genetically modified yeast of the species *Issatchenkia orientalis* and closely relates species, and fermentation processes using same. US patent 8,097,4482012.
29. Park HJ, Bae JH, Ko HJ, et al. Low-pH production of d-lactic acid using newly isolated acid tolerant yeast *Pichia kudriavzevii* NG7. *Biotechnol Bioeng*. 2018;115(9):2232-2242.
30. Seo SH, Rhee CH, Park HD. Degradation of malic acid by *Issatchenkia orientalis* KMBL 5774, an acidophilic yeast strain isolated from Korean grape wine pomace. *J Microbiol*. 2007;45(6):521-527.
31. Suthers PF, Dinh HV, Fatma Z, et al. Genome-scale metabolic reconstruction of the non-model yeast *Issatchenkia orientalis* SD108 and its application to organic acids production. *Metab Eng Commun*. 2020; 11:e00148.
32. Thalagala TATP, Kodama S, Mishima T, et al. Study on ethanol fermentation using D-glucose rich fractions obtained from lignocelluloses by a two-step extraction with sulfuric acid and *Issatchenkia orientalis* MF 121. *J Appl Glycosci*. 2009;56(1):7-11.
33. Tran VG, Cao M, Fatma Z, Song X, Zhao H. Development of a CRISPR/Cas9-based tool for gene deletion in *Issatchenkia orientalis*. *mSphere*. 2019;4(3).
34. Cao M, Fatma Z, Song X, et al. A genetic toolbox for metabolic engineering of *Issatchenkia orientalis*. *Metab Eng*. 2020;59:87-97.
35. Norsigian CJ, Pusarla N, McConn JL, et al. BiGG models 2020: multi-strain genome-scale models and expansion across the phylogenetic tree. *Nucleic Acids Res*. 2020;48(D1):D402-D406.
36. Moretti S, Martin O, Van Du TT, Bridge A, Morgat A, Pagni M. MetaNetX/MNXref-reconciliation of metabolites and biochemical reactions to bring together genome-scale metabolic networks. *Nucleic Acids Res*. 2016;44(D1):D523-D526.
37. Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, Stevens RL. High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat Biotechnol*. 2010;28(9):977-982.
38. Hastings J, Owen G, Dekker A, et al. ChEBI in 2016: improved services and an expanding collection of metabolites. *Nucleic Acids Res*. 2016;44(D1):D1214-D1219.
39. Kim S, Chen J, Cheng T, et al. PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res*. 2019;47(D1):D1102-D1109.

40. Burgard AP, Nikolaev EV, Schilling CH, Maranas CD. Flux coupling analysis of genome-scale metabolic network reconstructions. *Genome Res.* 2004;14(2):301-312.
41. Mahadevan R, Schilling CH. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab Eng.* 2003;5(4):264-276.
42. Courtot M, Juty N, Knupfer C, et al. Controlled vocabularies and semantics in systems biology. *Mol Syst Biol.* 2011;7:543.
43. Patil KR, Rocha I, Forster J, Nielsen J. Evolutionary programming as a platform for in silico metabolic engineering. *BMC Bioinf.* 2005; 6:308.
44. Ebrahim A, Lerman JA, Palsson BO, Hyduke DR. COBRApy: COConstraints-based reconstruction and analysis for python. *BMC Syst Biol.* 2013;7:74.
45. McKinney W. Data Structures for Statistical Computing in Python. Paper presented at: Proceedings of the 9th Python in Science Conference; 2010.
46. Feng S, Heath E, Jefferson B, et al. Hypergraph models of biological networks to identify genes critical to pathogenic viral response. *BMC Bioinf.* 2021;22(1):287.
47. Jørgensen H, Kristensen JB, Felby C. Enzymatic conversion of lignocellulose into fermentable sugars: challenges and opportunities. *Biofuels Bioprod Biorefin.* 2007;1(2):119-134.
48. Rashid T, Kait CF, Murugesan T. Conversion of lignocellulose into fermentable sugars using solid acid catalysis – a review. *Appl Mech Mater.* 2014;625:341-344.
49. Swiatek K, Gaag S, Klier A, Kruse A, Sauer J, Steinbach D. Acid hydrolysis of lignocellulosic biomass: sugars and furfurals formation. *Catalysts.* 2020;10(4):437.
50. Werpy T, Petersen G. Top value added chemicals from biomass, vol. I: Results of screening for potential candidates from sugars and synthesis gas. 2004.
51. Zanchi NE, Gerlinger-Romero F, Guimaraes-Ferreira L, et al. HMB supplementation: clinical and athletic performance-related effects and mechanisms of action. *Amino Acids.* 2011;40(4):1015-1025.
52. Walther T, Topham CM, Irague R, et al. Construction of a synthetic metabolic pathway for biosynthesis of the non-natural methionine precursor 2,4-dihydroxybutyric acid. *Nat Commun.* 2017;8:15828.
53. Campodonico MA, Andrews BA, Asenjo JA, Palsson BO, Feist AM. Generation of an atlas for commodity chemical production in *Escherichia coli* and a novel pathway prediction algorithm, GEM-Path. *Metab Eng.* 2014;25:140-158.
54. Dinh HV, King ZA, Palsson BO, Feist AM. Identification of growth-coupled production strains considering protein costs and kinetic variability. *Metab Eng Commun.* 2018;7:e00080.
55. Kumar V, Ashok S, Park S. Recent advances in biological production of 3-hydroxypropionic acid. *Biotechnol Adv.* 2013;31(6):945-961.
56. Tan MJ, Chen X, Wang YK, Liu GL, Chi ZM. Enhanced citric acid production by a yeast *Yarrowia lipolytica* over-expressing a pyruvate carboxylase gene. *Bioprocess Biosyst Eng.* 2016;39(8): 1289-1296.
57. Zou X, Guo L, Huang L, et al. Pathway construction and metabolic engineering for fermentative production of beta-alanine in *Escherichia coli*. *Appl Microbiol Biotechnol.* 2020;104(6):2545-2559.
58. Raab AM, Gebhardt G, Bolotina N, Weuster-Botz D, Lang C. Metabolic engineering of *Saccharomyces cerevisiae* for the biotechnological production of succinic acid. *Metab Eng.* 2010;12(6):518-525.
59. Zahoor A, Kuttner FTF, Blank LM, Ebert BE. Evaluation of pyruvate decarboxylase-negative *Saccharomyces cerevisiae* strains for the production of succinic acid. *Eng Life Sci.* 2019;19(10):711-720.
60. Kim SR, Kwee NR, Kim H, Jin YS. Feasibility of xylose fermentation by engineered *Saccharomyces cerevisiae* overexpressing endogenous aldose reductase (GRE3), xylitol dehydrogenase (XYL2), and xylulokinase (XYL3) from *Scheffersomyces stipitis*. *FEMS Yeast Res.* 2013;13(3):312-321.
61. Ranganathan S, Suthers PF, Maranas CD. OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. *PLoS Comput Biol.* 2010;6(4):e1000744.
62. King ZA, O'Brien EJ, Feist AM, Palsson BO. Literature mining supports a next-generation modeling approach to predict cellular byproduct secretion. *Metab Eng.* 2017;39:220-227.
63. van Maris AJ, Konings WN, van Dijken JP, Pronk JT. Microbial export of lactic and 3-hydroxypropanoic acid: implications for industrial fermentation processes. *Metab Eng.* 2004;6(4):245-255.
64. Thiele I, Palsson BO. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc.* 2010;5(1):93-121.
65. Andersen MR, Nielsen ML, Nielsen J. Metabolic model integration of the bibliome, genome, metabolome and reactome of *Aspergillus niger*. *Mol Syst Biol.* 2008;4:178.
66. Oftadeh O, Salvy P, Masid M, Curvat M, Miskovic L, Hatzimanikatis V. A genome-scale metabolic model of *Saccharomyces cerevisiae* that integrates expression constraints and reaction thermodynamics. *Nat Commun.* 2021;12(1):4790.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Suthers PF, Maranas CD. Examining organic acid production potential and growth-coupled strategies in *Issatchenkia orientalis* using constraint-based modeling. *Biotechnol. Prog.* 2022;38(5):e3276. doi:[10.1002/btpr.3276](https://doi.org/10.1002/btpr.3276)